On Nonrandomized, Randomized, and Fuzzy *p*-Values in Multiple Hypothesis Testing: A Unified Approach

Joshua D. Habiger
Oklahoma State University
jhabige@okstate.edu

Aug. 5, 2013

Example $X \sim Bin(11, p)$

1. Model: $X \sim Bin(11, p)$
2. Hypotheses: $H_0 : p = 1/2$ vs. $H_1 : p > 1/2$
3. Data: $X = 8$
4. p-value:
    - Conventional p-value = $\Pr_X(X \geq 8) = 0.11$
    - Mid p-value = $\Pr_X(X > 8) + 1/2 \Pr_X(X = 8) = 0.072$ [1]
5. Conclusion: Fail to reject H_0 at $\alpha = 0.05$.

[1] Lancaster (1961)

Reconciliation

Client: "But the mid-p-value is almost 0.05!!!"

Statistician: "Well the type 1 error rate for mid-p-value-based decision $\approx$ 0.05"

Proposed Statement: "The mid-p-value is 0.072 and is conservative for a level $\alpha = 0.05$ decision rule"[2]

[2]Note: Actual type 1 error rate = $\Pr_X^0(X \geq 9) = 0.033$

Goals

Main Points:

1. Communicate the behavior (liberal/conservative/etc.) of decision rule - especially in borderline cases
    - Behavior understood via test function $\phi(x)$
2. How can we compute $\phi(x)$ in complicated settings, ex. multiple hypothesis testing?

Test Function

- Let X have countable support and consider H_0 vs. H_1

- A **Size α test function** $\phi(x;\alpha)$
  - $\phi(x;\alpha) \in [0,1]$
  - $E_X[\phi(X;\alpha)] = \alpha$ under H_0

-Example

$$\phi^*(x;\alpha) = \begin{cases} 1 & x > k(\alpha) \\ \gamma(\alpha) & x = k(\alpha) \\ 0 & x < k(\alpha) \end{cases}$$

Decision function and p-value

- If $\phi(x;\alpha) \in (0,1)$ make decision using $u \in (0,1]$
  - Decision function: $\delta(x,u;\alpha) = I(u \leq \phi(x;\alpha)) \in \{0,1\}$
  - p-value: $p(x,u) = \inf\{\alpha : \delta(x,u;\alpha) = 1\}$[3]
- Example: $X \sim Bin(11,p)$ with

$$\phi^*(x;0.05) = \begin{cases} 1 & x > 8 \\ 0.21 & x = 8 \\ 0 & x < 8 \end{cases}$$

  - $\delta^*(8,u;0.05) = I(u \leq 0.21)$
  - $p^*(8,u) = \Pr_X(X > 8) + u \Pr_X(X = 8) = 0.033 + u0.081$

[3] Habiger and Peña(2011), Peña, Habiger, Wu (2012)

How to produce u?

- $\delta(x, u; \alpha)$ and $p(x, u)$ are
  - **nonrandomized** if u **chosen**
    - Mid-p-value $p(x, 1/2)$
    - Conventional p-value $p(x, 1)$
  - **randomized** if u **generated**

- $p(x, U)$ is an **abstract randomized** (fuzzy) p-value[4] if $U \sim Un(0, 1)$
  - Example:
    $p^*(8, U) = 0.033 + U0.081 \sim Un(0.033, 0.114)$

[4] Geyer and Meeden (2005)

Proposed Method

Step 1 **Compute $\phi(x;\alpha)$. Reject or fail to reject H_0 if possible and stop. Else report $p(x,U)$ and $\phi(x;\alpha)$ and go to Step 2a or Step 2b.**

Step 2a **Generate u**, compute $\delta(x,u;\alpha)$ and $p(x,u)$

Step 2b **Specify u**, compute $\delta(x,u;\alpha)$ and $p(x,u)$

- *Usual Approach: Go directly to Step 2a or Step 2b*

- Viewpoint[5]: Goal to "estimate" $\phi(x;\alpha) \in [0,1]$ with $\delta(x,u;\alpha) \in \{0,1\}$

[5] Blyth and Staudte (1995)

Step 1 vs. 2a

Q: What if we only report $\delta(x, u; \alpha)$ but not $\phi(x; \alpha)$ in Step 2a?

Mathematical Answer:

Theorem

Let U be uniformly distributed over $[0, 1]$ and independent of X, then the following claims are true:

C1: $E_U(\delta(x, U; \alpha)) = \phi(x; \alpha)$ *and hence* $E_{(X,U)}[\delta(X, U; \alpha)] = E_X[\phi(X; \alpha)]$ *(unbiased),*

C2: $Var(\delta(X, U; \alpha)) \geq Var(\phi(X; \alpha))$.

Intuitive Answer: Information loss

- Did $\delta(x, u; \alpha)$ depend on u?

Step 1 vs. 2b

Q: What if we only report $\delta(x, u; \alpha)$ but not $\phi(x; \alpha)$ in Step 2a?

Mathematical Answer:

Theorem

For any fixed or specified value of u,

C3: $E_X[\delta^*(X, u; \alpha)] \neq E_X[\phi^*(X; \alpha)]$ *(biased) for every* $\gamma(\alpha) \in (0, 1)$. *In particular,* $E_X[\delta^*(X, u; \alpha)] > (<) E_X[\phi^*(X; \alpha)]$ *for* $\gamma(\alpha) < (\geq) u$, *ex.* $u > \gamma \Rightarrow$ *decision conservative (size* $< \alpha$*)*

Intuitive Answer: Information loss

- Did $\delta(x, u; \alpha)$ depend on u?
- Is $\delta(x, u; \alpha)$ conservative or liberal?

Example: $u = 1/2 > 0.21 = \phi^*(8; 0.05) \Rightarrow$ size $< \alpha$.

Multiple decision function

Brief overview:

- Goal: Test H_{0m}, $m = 1, 2, ..., M$ null hypotheses with data $\mathbf{x} = (x_1, x_2, ..., x_M)$ and $\mathbf{u} = (u_1, u_2, ..., u_M)$
- p-values $\mathbf{p} = \mathbf{p}(\mathbf{x}, \mathbf{u}) = [p_1(x_1, u_1), p_2(x_2, u_2), ..., p_M(x_M, u_M)]$
- Multiple Testing procedure is a multiple decision function (MDF)

$$\boldsymbol{\delta}(\mathbf{x}, \mathbf{u}; \alpha) = [\delta_1(\mathbf{x}, \mathbf{u}; \alpha), \delta_2(\mathbf{x}, \mathbf{u}; \alpha), ..., \delta_M(\mathbf{x}, \mathbf{u}; \alpha)] \in \{0, 1\}^M$$

where $\delta_m(\mathbf{x}, \mathbf{u}; \alpha) = \delta_m(\mathbf{p}(\mathbf{x}, \mathbf{u}); \alpha)$

Example: Benjamini and Hochberg (1995)

1. $k(\mathbf{p}) = \max\left\{i : p_{(i)} \leq \alpha \frac{i}{M}\right\}$
2. $\delta_m^{BH}(\mathbf{x}, \mathbf{u}; \alpha) = I\left(p_m(x_m, u_m) \leq \alpha \frac{k(\mathbf{p})}{M}\right)$

Adjusted *p*-values

Adjusted *p*-value for MTP:

$$q_m(\mathbf{x}, \mathbf{u}) = \inf\{\alpha : \delta_m(\mathbf{x}, \mathbf{u}; \alpha) = 1\}$$

- **Adjusted nonrandomized** *p*-value $q_m(\mathbf{x}, \mathbf{u})$: $\mathbf{u}$ is chosen
- **Adjusted randomized** *p*-value $q_m(\mathbf{x}, \mathbf{u})$: $\mathbf{u}$ generated
- **Adjusted abstract randomized** (fuzzy) *p*-value $q_m(\mathbf{x}, \mathbf{U})$
  - We can easily sample from $q_m(\mathbf{x}, \mathbf{U})$ via $q_m(\mathbf{x}, \mathbf{u}^1), q_m(\mathbf{x}, \mathbf{u}^2), \ldots$ and construct a histogram

Remark: $q_m(\mathbf{x}, \mathbf{u})$ can often be computed with software

Multiple Test Function

Idea: Recall $\phi(x;\alpha) = E_U[\delta(x, U;\alpha)]$.

Definition

Define multiple test function

$$\boldsymbol{\phi}(\mathbf{x};\alpha) = [\phi_1(\mathbf{x};\alpha), \phi_2(\mathbf{x};\alpha), ..., \phi_M(\mathbf{x};\alpha)] \in [0,1]^M$$

where

$$\phi_m(\mathbf{x};\alpha) = E_{\mathbf{U}}[\delta_m(\mathbf{x},\mathbf{U};\alpha)] = \int_0^1\int_0^1 ... \int_0^1 \delta_m(\mathbf{x},\mathbf{u};\alpha)du_1du_2...du_M$$

for $m = 1, 2, ..., M$.

Remark: each $\phi_m(\mathbf{x};\alpha)$ can be easily computed numerically.

Microarray Example

Table: A portion of the microarray data in Timmons et. al (2007).

gene	brown fat cell				white fat cell			
m	$x_{m,1}$	$x_{m,2}$	...	$x_{m,10}$	$y_{m,1}$	$y_{m,2}$	...	$y_{m,14}$
1	1.22	1.66	...	1.41	5.64	1.79	...	11.50
2	3.57	19.22	...	5.23	5.17	29.49	...	7.58
.	.	.	...	.	.	.	...	.
.	.	.	...	.	.	.	...	.
M=12488	2.52	10.91	...	22.67	10.70	7.35	...	21.95

1. Compute shifted Wilcoxon rank sum stat: $w_m = \left| w_m^* - \frac{14x10}{2} \right|$
2. Compute $p_m^*(w_m, u_m) = \Pr_W(W_m > w_m) + u_m \Pr_W(W_m = w_m)$
3. Apply MTP: Storey (2002, 2004) adaptive FDR procedure using $q.value()$ with $\alpha = 0.05$.

Step 2a

- What if skip step 1, generate ***u***, and apply MTP

Step 1	$\phi_m(\mathbf{x}; 0.05)$	0	1	????
	Count	????	????	????
			↙	
Step 2a	Count	9315	3173	

Step 1 + 2a

- What information did Step 1 provide?

Step 1	$\phi_m(\mathbf{x}; 0.05)$	0	1	**0.94**
	Count	9302	3033	**153**
Step 2a	Count	9315	3173	↙

Step 1 tells us . . .

- 153 decisions made randomly!
- 140 genes “discovered” randomly!

Step 2b

- What if we **skip Step 1**, choose $u = 1/2$, and apply MTP

Step 1	$\phi_m(\mathbf{x}; 0.05)$	0	1	????
	Count	????	????	????
			↙	
Step 2b	Count	9302	3186	

Step 1 + 2b

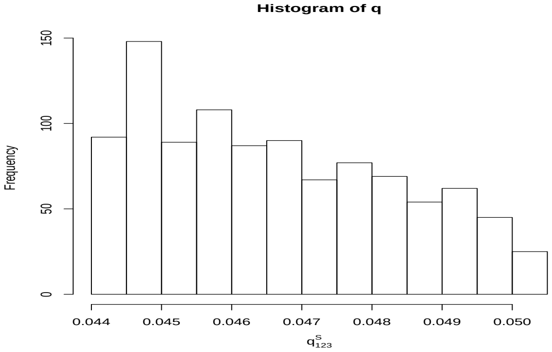
- What information did Step 1 provide?

Step 1	$\phi_m(\mathbf{x}; 0.05)$	0	1	**0.94**
	Count	9302	3033	**153**
				↙
Step 2b	Count	9302	3186	

Step 1 tells us . . .

- 153 genes "discovered" because $u = 1/2 < 0.94$
- **Procedure Liberal**

Adjust Fuzzy *p*-value



- $0.044 \leq q_m(\mathbf{x}, \mathbf{U}) \leq 0.051$
- Observe Theorem: $\Pr_{\mathbf{U}}(q_m(\mathbf{x}, \mathbf{U}) \leq \alpha) = \phi_m(\mathbf{x}) = 0.94$

Main Point

Main Point

1. Like it or not must specify or generate u to make some decisions
2. We should tell our clients when decisions were made with u and report liberal/conservative/etc behavior
   - $\phi(x)$ and $p(x, U)$ useful here

In Practice

1. Not simpler but . . . "as simple as possible?"

Loose ends

- When supports of test statistics equal
  - ~~$u_1 = u_2 = ...u_M = u = 1/2$~~
  - $u_1 = \frac{1}{M+1}, u_2 = \frac{2}{M+1}, \ldots, u_M = \frac{M}{M+1}$
- When supports of test statistics not equal - Tarone (1990)

$$\begin{array}{c}\text{Step 0: Automatically accept some } H_{0m}\text{s} \\ \Downarrow \\ \text{Step 1: Applied to remaining } H_{0m}\text{s as usual} \\ \Downarrow \\ \text{Step 2: Applied to remaining } H_{0m}\text{s as usual}\end{array}$$

References

Agresti, A. and A. Gottard (2007). Nonconservative exact small-sample inference for discrete data. Computational Statistics & Data Analysis 51(12), 6447–6458.

Barnard, G. A. (1989). On the alleged gains in power from lower *p*-values. Statistics in Medicine 8(12), 1469 – 1477.

Blyth, C. R. and R. G. Staudte (1995). Estimating statistical hypotheses. Statistics and Probability Letters 23(1), 45 – 52.

Geyer, C. J. and G. D. Meeden (2005). Fuzzy and randomized confidence intervals and *P*-values. Statistical Science 20(4), 358–387. With comments and a rejoinder by the authors.

Habiger, J. and E. Peña (2011). Randomized p-values and nonparametric procedures in multiple testing. Journal of Nonparametric Statistics 23(3), 583–604.

Kulinskaya, E. and A. Lewin (2009). On fuzzy familywise error rate and false discovery rate procedures for discrete distributions. Biometrika 96(1), 201–211.

Lancaster, H. O. (1961). Significance tests in discrete distributions. Journal of the American Statistical Association 56, 223–234.

Peña, E., J. Habiger, and W. Wu (2011). Power-enhanced multiple decision functions controlling family-wise error and false discovery rates. The Annals of Statistics 39(1), 556 – 583.

Storey, J. (2002). A direct approach to false discovery rates. Journal of the Royal Statistical Society. Series B 64(3), 479 – 498.

Storey, J. D., J. E. Taylor, and D. Siegmund (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. Journal of the Royal Statistical Society. Series B. 66(1), 187–205.

Tarone, R. E. (1990). A modified bonferroni method for discrete data. Biometrics 46(2), 515–522.

Timmons, J., K. Wennmalm, O. Larsson, T. Walden, T. Lassmann, N. Petrovic, D. Hamilton, R. Gimeno, C. Wahlestedt, K. Baar, J. Nedergaard, and B. Cannon (2007). Myogenic gene expresion signature establishes that brown and white adipocytes originate from distinct cell lineages. Proceedings of the National Academy of Sciences of the United States of America 104 (11), 4401 – 4406.